Published in final edited form as:
J Am Stat Assoc. 2014 September 1; 109(507): 1174–1187. doi:10.1080/01621459.2014.881743.

Targeted Local Support Vector Machine for Age-Dependent Classification

Tianle Chen [graduate student],
Department of Biostatistics, Mailman School of Public Health, Columbia University

Yuanjia Wang [Associate Professor],
Department of Biostatistics, Mailman School of Public Health, Columbia University

Huaihou Chen [post-doctoral fellow],
New York University

Karen Marder [Professor], and
Department of Neurology, Psychiatry (at the Sergievsky Center and Taub Institute), Columbia University Medical Center

Donglin Zeng[1] **[Professor]**
Department of Biostatistics, University of North Carolina at Chapel Hill

Abstract

We develop methods to accurately predict whether pre-symptomatic individuals are at risk of a disease based on their various marker profiles, which offers an opportunity for early intervention well before definitive clinical diagnosis. For many diseases, existing clinical literature may suggest the risk of disease varies with some markers of biological and etiological importance, for example age. To identify effective prediction rules using nonparametric decision functions, standard statistical learning approaches treat markers with clear biological importance (e.g., age) and other markers without prior knowledge on disease etiology interchangeably as input variables. Therefore, these approaches may be inadequate in singling out and preserving the effects from the biologically important variables, especially in the presence of potential noise markers. Using age as an example of a salient marker to receive special care in the analysis, we propose a local smoothing large margin classifier implemented with support vector machine (SVM) to construct effective age-dependent classification rules. The method adaptively adjusts age effect and separately tunes age and other markers to achieve optimal performance. We derive the asymptotic risk bound of the local smoothing SVM, and perform extensive simulation studies to compare with standard approaches. We apply the proposed method to two studies of premanifest Huntington's disease (HD) subjects and controls to construct age-sensitive predictive scores for the risk of HD and risk of receiving HD diagnosis during the study period.

[1]*Address for correspondence*: Yuanjia Wang, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032. yuanjia.wang@columbia.edu.

Supplementary Material: The online supplementary material contains the proof of Theorem 1.

Keywords

Statistical learning; Local smoothing; Reproducing kernel Hilbert space; Risk bound; Huntington's disease

1 Introduction

An important research goal for chronic diseases is to develop effective early intervention to delay onset, slow disease progression, and provide different treatment or care management at each stage based on subject-specific characteristics (Paulsen et al., 2006). It is necessary to identify biological, behavioral and clinical markers that can be combined to distinguish premanifest subjects at high risk of a disease from those who are at low risk or free of risk. For many illnesses, existing clinical literature may suggest the risk of disease varies with some markers of biological and etiological importance. For example, it is well known that the risk of Alzheimer's disease increases with age (Celsis, 2000), and the predictive power of other markers and their relative importance often change over a subject's lifespan. It is beneficial to take advantage of the existing etiologic information on disease risk to develop age-sensitive diagnostic rules in conjunction with other markers with less clear prior biological information on disease risk to boost predictive power. Using age as an example of a salient marker to receive special care in the analysis, we develop methods to treat biologically important variables separately from other variables in the presence of some potential noise markers. The developed prediction rules have implications on prioritizing other markers and informing timing of therapeutic interventions to guide personalized medicine.

To predict binary outcomes such as disease status, regression-based methods including logistic regression and time-varying coefficient models are often used (Cai et al., 2000; Wang et al., 2009b). These models focus on estimating population-average association (e.g., odds ratio) instead of making subject-specific prediction or classification, thus may not be optimal (Pepe et al., 2004, 2006; Ware, 2006). For example, variables that are themselves not significant at certain levels may contribute to improving prediction in combination especially when they are highly correlated (Wei et al., 2009). To directly focus on classification and prediction, large margin-based statistical learning approaches (e.g., Vapnik, 1995; Shen et al., 2003; Zhang et al., 2006; Wang et al., 2009a; Wu and Liu, 2012) can be used. The geometric set up of these methods is to construct an optimal separating boundary between two classes by maximizing the margin from each class to the boundary. The equivalent statistical framework is to minimize a margin-based loss function subject to a regularization penalty. They are among the most successful nonparametric and robust classifiers in practice that can improve individual-specific prediction and classification problems especially in high-dimensional settings with correlated variables (Moguerza and Munoz, 2006; Orru et al., 2012). Among the large-margin based classifiers, support vector machine (SVM) is one of the most popular binary classifiers proven to exhibit some optimal theoretical properties (Lin, 2002). Recently, Ladicky and Torr (2011) and Zhang et al. (2011b) proposed a non-specific locally linear smoothing in the SVM context using all the features variables. However, what they considered is based on local affine approximation of

the entire feature variable space involving variables in all dimensions. Their locality is defined by all the features variables in a neighborhood of a data point. When the dimension of the feature variable space is high, it may be difficult to perform smoothing in the entire feature space due to sparseness of data in any local neighborhood. In addition, since these approaches are based on linearization of a potentially high-dimensional nonparametric surface, stronger assumptions on the smoothness of separating boundary in all dimensions of the feature space are required.

One convenient approach to incorporate age information to classify a subject's at-risk status is to treat age as one of input variables interchangeably with other markers and learn classification rules using kernel machine (e.g., Gaussian kernel). However, such a strategy may not be optimal for several reasons. First, from a clinical point of view, age plays distinctive clinical and biological roles on disease risk. It is the easiest factor to measure to be used for indicating the timing of intervention and guiding choices of treatments, and thus it should call for some special attention. Second, from a statistical point of view, lumping age together with other markers exchangeably in a learning algorithm is very likely to dilute the age signal especially when the marker dimension is not small and some noise variables are included. Furthermore, since all variables are tuned by the same tuning parameter, age effect may be masked by the other markers which potentially introduce noise. This is observed in our subsequent numerical studies. Lastly, using fully nonparametric learning without distinguishing age from other markers makes it difficult to provide an interpretable and practical guideline for timely intervention.

In this work, we develop a large-margin based classifier implemented with SVM for discriminating subjects at risk through solving a kernel weighted optimization problem to provide age-dependent prediction rules from markers collected in cross-sectional studies. Since disease risk for two subjects close in age is expected to be similar controlling for other characteristics, certain smoothness with respect to age is anticipated so it can be taken advantage of when classifying a subject's disease status. The proposed approach uses a local smoothing kernel to pool information across subjects similar in age and selects the tuning parameter for age separately from tuning parameter for other markers. Therefore, we adaptively estimate age effect and protect the age signal from being lost especially when noise markers are present. We first consider interpretable locally linear prediction rules where the age profile for each marker can be easily presented and used to assess importance of each marker. We then consider more general nonlinear prediction rules through kernel machines locally at each age. Our method differs from the literature (Ladicky and Torr, 2011; Zhang et al., 2011b) in that there exists a targeted variable with strong prior knowledge to be predictive or needs to be adjusted. We perform local smoothing along one targeted dimension of a well-motivated content-important variable (e.g., age) while leaving other variables intact. Our approach only requires data to be reasonably abundant along one targeted dimension.

The remainder of work is organized as follows. In section 2, we describe the details of the proposed method and provide an easy computational algorithm supporting the method. In section 3, we study the theoretical properties of the risk bound as a function of local smoothing kernel bandwidth. In section 4, we perform extensive simulation studies to

compare the proposed method with several alternative approaches and examine the finite sample properties of the fitted classification boundaries. In section 5, we apply the proposed methods to two Huntington's disease (HD) data examples (Dorsey and Huntington Study Group COHORT Investigators, 2012; Paulsen et al., 2008) to predict age-specific risk of developing HD or risk of pre-symptomatic subjects receiving HD diagnosis during study period using motor, cognitive and behavioral markers, and show the age-dependent profiles of several key markers. Some concluding remarks are given in section 6.

2 Targeted local smoothing for large margin classifiers

Let D be the dichotomous at-risk status coded as 1 and −1 for subjects at risk of a disease and not at risk, respectively. Let W denote a subject's age and let $\boldsymbol{X}$ be a vector of the other potential risk-altering markers for this subject. The goal is to determine an age-dependent classification rule using $\boldsymbol{X}$ to predict D at each age W (the target variable). For this purpose, we first consider the following composite predictive score

$$\alpha(W)+\boldsymbol{X}^T\boldsymbol{\beta}(W), \quad (1)$$

where $\alpha(W)$ is an unspecified baseline function, and $\boldsymbol{\beta}(W)$ is a vector of unspecified age-dependent coefficients for markers $\boldsymbol{X}$. A subject with a positive fitted score will be classified as at risk of disease, and as risk free if the subject has a negative fitted score. Note the score in model (1) has a nonparametric form with respect to age effect, while at each given age it is linear in terms of markers $\boldsymbol{X}$. This formulation allows decomposition of the diagnostic score as the sum of a component due to normal aging, $\alpha(W)$, and a component due to the other markers, $\boldsymbol{X}^T\boldsymbol{\beta}(W)$. The unrestricted form of $\boldsymbol{\beta}(w)$ allows the age-dependent effect to change freely. Since age may serve as a surrogate for many unmeasured physiological factors, for subjects close in age and with the same values of other markers, the disease risk is expected to be similar, and thus certain smoothness is expected for functions $\alpha(w)$ and $\boldsymbol{\beta}(w)$.

The age-dependent classification boundary in (1) has several features. First, although the score is allowed to change from one age to another in an unspecified fashion, at a given age the prediction is a linear combination of markers to facilitate interpretation. It is easy to tell which markers are effective at which age by examining coefficient functions $\boldsymbol{\beta}(w)$. When varying age smoothly, the corresponding separating hyperplane constructed from other markers also changes smoothly. Second, since the coefficient function $\boldsymbol{\beta}(w)$ is age-adaptive, it captures the age-dependent effects of markers. As introduced later in section 5, there might be markers informative for younger subjects but not older subjects or vice versa, which suggests different sets of markers would be considered as effective depending on a subject's age. Third, some cumulative summary of $\boldsymbol{\beta}(w)$, for example, the vector $\int|\boldsymbol{\beta}(w)|dw$, can be used to rank the overall importance of markers under model (1).

In a standard classification problem with predictive score $\alpha + \boldsymbol{X}^T\boldsymbol{\beta}$, a large-margin based classifier would minimize a penalized loss function,

$$\min_{\alpha,\boldsymbol{\beta}} \sum_i \mathscr{L}\{D_i, \boldsymbol{X}_i; \alpha, \boldsymbol{\beta}\} + \lambda_n \|\boldsymbol{\beta}\|^2,$$

where λ_n is a tuning parameter depending on the sample size, and $\mathscr{L}(\cdot)$ belongs to a class of margin-based loss functions. Examples of margin-based loss functions include hinge loss, i.e., SVM loss, $\{1 - df(\boldsymbol{x})\}_+$; its variations such as ψ−loss which satisfies $U \geq \psi(z) > 0$ where $z = df(\boldsymbol{x})$, if $z \in [0, \tau]$; $\psi(z) = U(1 - \text{sign}(z))$, otherwise for some constants U and $0 < \tau < 1$ (Shen et al., 2003); and logistic loss, $\log\{1 + \exp(-df(\boldsymbol{x}))\}$. To fit the age-dependent predictive score in model (1) taking advantage of the smoothness effect in age, we introduce a local smoothing kernel weighted support vector machine (KSVM). Essentially, the KSVM solves an SVM at each w_0 where the ith subject is weighted by a local smoothing kernel function $K_{h_n}(W_i - w_0)$, so we pool information across subjects whose ages are close to w_0. Here $K_{h_n}(\cdot)$ is a symmetric kernel density and h_n is its bandwidth. Specifically, we fit (1) by solving

$$\min_{\alpha(w_0),\boldsymbol{\beta}(w_0)} \sum_i K_{h_n}(W_i - w_0) \mathscr{L}\{D_i, \boldsymbol{X}_i; \alpha(w_0), \boldsymbol{\beta}(w_0)\} + \lambda_n \|\boldsymbol{\beta}(w_0)\|^2, \quad (2)$$

where w_0 varies across the support of age W_i. The loss function in the minimization problem (2) can be considered as a locally weighted loss where the subjects closer to age w_0 contribute larger weights.

In the subsequent implementation of KSVM, we choose the hinge loss. Computationally, the optimization problem is solved by

$$\min_{\alpha(w_0),\boldsymbol{\beta}(w_0)} \sum_i K_{h_n}(W_i - w_0)\xi_i + \lambda_n \|\boldsymbol{\beta}(w_0)\|^2, \text{ subject to } D_i\{\alpha(w_0) + \boldsymbol{X}_i^T \boldsymbol{\beta}(w_0)\} \geq 1 - \xi_i, \xi_i \geq 0.$$

This alternative form provides some insights to the locally weighted objective function (2). Treating the slack variables ξ_i as serving similar roles as residuals in a regression model, problem (2) can be thought as minimizing a penalized locally weighted “residual” subject to linear constraints. Using the Lagrange multipliers, we can derive the corresponding dual form as

$$\max_{\gamma \in \mathbb{R}^n} \sum_i \gamma_i - \frac{1}{2} \sum_{i,j} \gamma_i \gamma_j D_i D_j \boldsymbol{X}_i^T \boldsymbol{X}_j, \text{ subject to } 0 \leq \gamma_i \leq K_{h_n}(W_i - w_0) C_n, \text{and} \sum_i \gamma_i D_i = 0.$$

Note that by reparametrizing γ_i as $\gamma_i K_{h_n}(W_i - w_0)$, the dual form is equivalent to

$$\begin{aligned} \max_{\gamma \in \mathbb{R}^n} \sum_i \gamma_i K_{h_n}(W_i - w_0) - \frac{1}{2} \sum_{i,j} \gamma_i \gamma_j D_i D_j K_{h_n}(W_i - w_0) K_{h_n}(W_j - w_0) \boldsymbol{X}_i^T \boldsymbol{X}_j, \\ \text{subject to } 0 \leq \gamma_i \leq C_n, \text{and} \sum_i \gamma_i K_{h_n}(W_i - w_0) D_i = 0. \end{aligned} \quad (3)$$

This is a locally weighted quadratic programming problem with linear constraints which can be solved conveniently using existing quadratic programming packages in R or MatLab. The resulting prediction of disease status for a *w*-year-old subject with markers $\boldsymbol{x}$ is

$$\hat{d}(\boldsymbol{x}, w)=\text{sign}\{\hat{f}(\boldsymbol{x};w)\}, \hat{f}(\boldsymbol{x};w)=\hat{\alpha}(w)+\boldsymbol{x}^T\hat{\beta}(w). \quad (4)$$

When at a given age the disease risk groups cannot be adequately separated by a linear function of marker, it may be useful to perform prediction in the reproducing kernel Hilbert space (RKHS, Wahba, 1990) feature space instead of the original marker space. Consider a nonparametric predictive score, $f(\boldsymbol{X}_i;w_i)$, which is a completely unspecified function of age and markers. The age-dependent decision boundary (1) corresponds to a special case of taking a linear combination of all components of $\boldsymbol{X}_i$ at each age point *w*, i.e., $\alpha(w_i)+\boldsymbol{X}_i^T\boldsymbol{\beta}(w_i)$. The nonlinear classification boundary relaxes the linear form (in terms of markers) at each age. To fit this nonlinear predictive score, we smooth age effect by a local smoothing kernel while mapping other markers to a RKHS feature space through Mercer kernels. To be specific, denote a Mercer kernel $H(\boldsymbol{x}, \boldsymbol{y})$, through an appropriate inner product in the RKHS. Commonly used Mercer kernels include the Gaussian kernel, where $H(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\gamma \|\boldsymbol{x} - \boldsymbol{y}\|^2)$, and the *k*th order polynomial kernel, where $H(\boldsymbol{x}, \boldsymbol{y}) = (1+\boldsymbol{x}^T \boldsymbol{y})^k$. At a given age w_0, the general decision boundary can be expressed as a function in the RKHS associated with $H(\cdot,\cdot)$ as

$$f(\boldsymbol{x};w_0)=\eta_0(w_0)+\sum_i \eta_i(w_0)H(\boldsymbol{X}_i, \boldsymbol{x}).$$

Comparing with the age-dependent model in (1), we see the methodology developed there can be implemented similarly. To pool information from subjects with similar ages, we use local smoother to weight observations around w_0, and the resulting local optimization problem is

$$\min_{\eta_0(w_0),\boldsymbol{\theta}(w_0)} \sum_i K_{h_n}(w_i-w_0)\{1 - D_i[\eta_0(w_0)+\sum_j \eta_j(w_0)H(\boldsymbol{X}_j, \boldsymbol{X}_i)]\}_+ + \lambda_n \|f(\cdot;w_0)\|^2_{\mathscr{H}},$$

where $\|f\|_{\mathscr{H}}$ is the norm of *f* in the RKHS. This locally weighted problem is solved in the dual space by replacing $\boldsymbol{X}_i^T\boldsymbol{X}_j$ in (3) by $H(\boldsymbol{X}_i, \boldsymbol{X}_j)$ associated the RKHS. The predicted at-risk status for a subject with marker $\boldsymbol{x}$ at age *w* using a fully nonparametric boundary is

$$\hat{d}(w, \boldsymbol{x})=\text{sign}\{\hat{f}(\boldsymbol{x};w)\}.$$

Note the distinct roles of the smoothing kernel $K_{h_n}(\cdot)$ and Mercer kernel $H(\cdot,\cdot)$: the former is used to pool information across age and the later for producing nonlinear decision boundary and dimension reduction with respect to the markers $\boldsymbol{X}$. The tuning parameters h_n and λ_n are chosen over a grid in a range, respectively, by minimizing the five-fold cross validated misclassification error. By using a different kernel and a separate tuning parameter for age,

the age effect can be better accommodated. In summary, the proposed method can be viewed as a splice of local smoothing and the RKHS framework for the SVM.

3 Theoretical results

In this section, we provide general theoretical results for the prediction errors using the fitted rule $f(\hat{\boldsymbol{X}}; w)$ as compared to the true optimal rule based on $f_0(\boldsymbol{X}; w) = 2P(D = 1|\boldsymbol{X}, W = w) - 1$. Our results require the following assumptions:

(C.1) Markers $(\boldsymbol{X}, W)$ have a bounded support and the conditional density of $(D, \boldsymbol{X})$ given $W = w$ and is twice-continuously differentiable with respect to w. Moreover, the marginal density for W is twice-continuously differentiable and bounded away from zero;

(C.2) The conditional distribution of $P(\boldsymbol{X}|W = w)$ has a uniform geometric noise exponent $\alpha > 0$; that is, there exists a constant C independent of w such that

$$\int |f_0(\boldsymbol{x};w)| \exp\left\{-\frac{\tau_{\boldsymbol{x}}(w)^2}{t}\right\} dP(\boldsymbol{x}|w) \leq Ct^{\alpha d/2},$$

where d is the dimension of $\boldsymbol{X}$, and $\tau_x(w)$ is the minimum distance from $\boldsymbol{x}$ to set $\{\boldsymbol{z} : f_0(\boldsymbol{z}; w) \leq 0\}$ for $\boldsymbol{x}$ with $f_0(\boldsymbol{x}; w) > 0$ while it is the minimum distance from $\boldsymbol{x}$ to set $\{\boldsymbol{z} : f_0(\boldsymbol{z}; w) \geq 0\}$ for $\boldsymbol{x}$ with $f_0(\boldsymbol{x}; w) < 0$;

(C.3) The kernel function $K_{h_n}(x) = h_n^{-1} K(x/h_n)$, where $K(\cdot)$ is symmetric and has finite second moments. The reproducing kernel Hilbert space used to fit the general decision boundary in (4) is generated from a Gaussian kernel with the bandwidth σ_n^{-1}.

(C.4) $h_n, \lambda_n \to 0$, $\sigma_n = \lambda_n^{-1/((\alpha+1)d)}$ and $\sqrt{n}h_n^2 \to \infty$.

Condition (C.1) ensures the smoothness of the distribution of the data over age W, so that we can borrow neighboring information to infer an age-dependent rule. Condition (C.2) is given in Steinwart and Scovel (2007), where they discussed a list of examples that satisfy the geometric noise exponent condition. In particular, if the distribution satisfies that $|f_0(\boldsymbol{x}; w)| \leq c\tau_x(w)^{\gamma 1}$ and $P(|f_0(\boldsymbol{x}; w)| \leq t|W = w) \leq Ct^q$ (Tsybakov noise exponent q), then condition (C.2) holds for $\alpha = (q+1)_{\gamma 1}/d$ if $q \geq 1$. In condition (C.4), as indicated in the proof and also in Steinwart and Scovel (2007), the choice of σ_n is optimal in terms of approximating the Bayesian error bound using the decision function for the reproducing kernel Hilbert space. Our main theoretical result is the following.

Theorem 1. Define $Err(f; w)$ as the prediction error at age w, i.e., $Err(f; w) = P(Df(\boldsymbol{X}; w) < 0|W = w)$. Under conditions (C.1)–(C.4), there exists a constant c_d such that for any $t > t_0$ where t_0 is a constant that depends on d, with probability at least $1 - e^{-t}$, it holds

$$\sup_{w \in \mathcal{W}} \left\{ \left| Err(\hat{f}; w) - Err(f_0; w) \right| \right\} \leq c_d(h_n^2 \lambda_n^{-1} + \lambda_n^{\alpha/(\alpha+1)} + r_n t),$$

where $r_n = n^{-1/2} h_n^{-2} \lambda_n^{-1-(d+2)/[(\alpha+1)d]}$ and is assumed to vanish as n goes to infinity.

Note the rate of risk bound is characterized through the geometric noise exponent α, local kernel smoothing parameter h_n, and the regularization parameter λ_n for SVM. In addition, we obtain the supreme norm risk bound over the support of age. The proof of Theorem 1 uses the embedding properties of the reproducing kernel Hilbert space, the large deviation results of empirical processes and the approximation using the kernel function. In the proof, we first note that $Err(\hat{f}; w) - Err(f_0; w)$ can be bounded by the corresponding risk based on the hinge loss $E[(1-D\hat{f})_+|W = w] - E[(1-Df_0)_+|W = w]$. We then decompose the latter into

$$E[(1-D\hat{f})_+|W = w] - E[(1-D\hat{f})_+ K_{h_n}(W - w)]/f_W(w) - \left\{E[(1-Df_0)_+|W=w] - E[(1-Df_0)_+ K_{h_n}(W-w)]/f_W(w)\right\}$$

and

$$\left\{E[(1-D\hat{f})_+ K_{h_n}(W-w)] - E[(1-Df_0)_+ K_{h_n}(W-w)]\right\}/f_W(w),$$

where f_W is the marginal density of W. Note that the first part is the bias due to the kernel smoothing so can be controlled using the kernel bandwidth. The latter part is a weighted version of the hinge loss; therefore, we will adapt the existing theory for the support vector machine (Steinwart and Scovel, 2007) but with careful modification due to the local smoothing kernel weights. The main challenge is to control the complexity of the kernel weighted functions from the reproducing kernel Hilbert space and assess the tail bound of some kernel weighted empirical processes. The detail of the proof is given in the Supplementary Material.

From Theorem 1, we conclude

$$\sup_{w \in \mathcal{W}} \left\{\left|Err(\hat{f}; w) - Err(f_0; w)\right|\right\} = O(h_n^2/\lambda_n + \lambda_n^{\alpha/(\alpha+1)}) + O_p(r_n).$$

Therefore, the optimal h_n is $[n^{-1/2}\lambda_n^{-1-(d+2)/[(\alpha+1)d]}]^{1/4}$ and the derived rate becomes

$$O_p(\lambda_n^{\alpha/(\alpha+1)} + [n^{-1/2}\lambda_n^{-1-(d+2)/[(\alpha+1)d]}]^{1/2}/\lambda_n).$$

This further gives the optimal choice of λ_n to be $\lambda_n^{opt} = n^{-\gamma}$ where $\gamma = 1/[6 + 2(d+2)/[(\alpha+1)d] + 4\alpha/(\alpha+1)]$ so it results in the optimal rate as

$$\sup_{w \in \mathcal{W}} \left\{\left|Err(\hat{f}; w) - Err(f_0; w)\right|\right\} = O_p(n^{-\gamma\alpha/(\alpha+1)}).$$

Clearly, these optimal rates depend on the unknown α, so they cannot be estimated. Instead, we suggest using the cross-validation to estimate the optimal choices of (h_n, λ_n) in practice. Under the special case when $f_0(\boldsymbol{x}; w) = \boldsymbol{X}^T\beta_0(w)$, if we choose $h_n^4/\lambda_n = n^{-1/2}$, then Theorem 1 can be modified to obtain

$$\sup_{w \in \mathcal{W}} |Err(\hat{f}; w) - Err(f_0; w)| = O_p(n^{-1/4}).$$

See the remark in the Supplementary Material. This rate gives an supreme bound of the classification error over the range of age when the underlying true classification boundary is linear.

4 Simulation studies

In this section, we conducted two sets of simulation studies to compare the empirical performance of KSVM with several alternatives. We generated samples with a size of n = 500 or 1000. For each setting we carried out 200 simulation runs. The standardized ages W_i were generated from a uniform distribution with support (0, 1). In the first set of experiments, we simulated data retrospectively. We generated dichotomous outcomes

$$Y_i = \text{sign}(W_i^2 + W_i - 1 + \varepsilon_i), \varepsilon_i \sim N(0, 1),$$

and given Y_i and W_i, we generated markers $\boldsymbol{X}_i = (X_{i1}, X_{i2})^T$ as

$$\boldsymbol{X}_i | Y_i, W_i \sim MVN\{\boldsymbol{\beta}(W_i) Y_i, \sigma^2 \boldsymbol{I}\},$$

and $\beta(w) = (\sin(4\pi w), 2\exp\{-20(w - 0.5)^2\})^T$.

We compared several alternative methods of handling age and other markers. For the handling of age effect we compared three approaches: (1) Using $\boldsymbol{X}_i$ but no W_i to train a standard SVM (SVM_0); (2) Using $\boldsymbol{X}_i$, W_i and $\boldsymbol{X}_i W_i$ as input variables to train a standard SVM (SVM_1); and (3) the proposed local smoothing SVM (KSVM). Within each of these methods, we compared using a linear kernel for input variables versus using a Gaussian kernel. To evaluate the performance of different approaches, we recorded the misclassification rate and area under the receiver operating characteristic (ROC) curve (AUC) at each age point, and computed an overall AUC and mean misclassification rate pooling data across all age points. For KSVM, the bandwidth h_n and the tuning parameter λ_n were chosen by 5-fold cross validation separately. For the multiple marker case, we included both markers X_{i1} and X_{i2}, and two other noise markers that do not contribute to disease risk.

Table 1 records the mean overall misclassification rate and AUC averaged over simulations for SVM_0, SVM_1 and KSVM with two choices of Mercer kernels when using X_{i1} alone, using X_{i2} alone, or using both plus two noise markers generated from a standard uniform distribution. From Table 1, when a single marker is used and the true classification boundary is more complex, such as a sine function, the locally weighted KSVM has much lower

average misclassification rate and much higher overall AUC than fitting SVM_0 or SVM_1. Using a Gaussian kernel improves overall performance for SVM_1 and KSVM but not for SVM_0. As expected, larger sample size improves AUC and decreases misclassification rate. When the underlying separating boundary is a simpler function such as a Gaussian function, the difference between KSVM and SVM_0 is still substantial while the difference between KSVM and SVM_1 is smaller. KSVM performs better than both SVM_0 and SVM_1 when a linear Mercer kernel is used. With a Gaussian kernel, the overall performance of SVM_1 and KSVM is similar due to the true coefficient function $\beta(w)$ being Gaussian (nonlinear) and the ability of Gaussian Mercer kernel to fit nonlinear separating boundaries. Furthermore, from this table, we observe that using all the markers compared to using single markers improves the prediction accuracy. In this case, comparing three approaches in treating the age effect, KSVM still has the overall performance superior to SVM_0 or SVM_1. The decrease in misclassification rate of KSVM over alternatives averaged across age and simulations is up to 50% (0.133 versus 0.265), and the increase in AUC is up to 13% (0.941 versus 0.817), which is substantial. Comparing different treatment of Mercer kernels, using a Gaussian kernel does not improve performance of either SVM_0, SVM_1 or KSVM.

Figure 1 presents more detailed information on the age-specific misclassification rate and AUC as a function of w averaged across simulation repetitions. When the true coefficient function is a sine function, KSVM dominates the alternatives over the entire range of w: it has lower misclassification rate and higher AUC at each age. For Gaussian coefficient function, KSVM improves upon SVM_0 and SVM_1 at the tail area. For the multiple marker case, Figure 1 (bottom panels) shows that while the age-specific AUC and misclassification rate indicates a superior performance of KSVM over the alternatives in the entire range of age, the improvement is much more significant at the tail area and at places where the two classes have large overlap. For example, SVM_1 fails to accommodate the decision boundary around about $w = 0.15$ and $w = 0.85$ (high misclassification rate and low AUC) as shown by two subfigures.

In the second set of simulations, we simulated data prospectively based on a known true decision boundary thus we could assess the performance of the fitted decision boundary through its mean squared error. First, we generated the standardized ages W_i from a uniform distribution with support (0, 1). The markers $\boldsymbol{X}_i = (X_{i1}, X_{i2})^T$ are generated as $\boldsymbol{\beta}(W_i) + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i$ follows $MVN(0, 1.5^2\boldsymbol{I})$ and the true $\boldsymbol{\beta}$'s are the same as in the first set of simulations. We further considered three different scenarios where $Z_{i1} = X_{i1} - \beta_1(W_i)$, $Z_{i2} = X_{i2} - \beta_2(W_i)$, and $Z_{i3} = Z_{i1} + Z_{i2}$, and the true margin had a width of $\delta = 0{:}3$. Then the class labels were generated as

$$Y_{ik} = \begin{cases} 1; \text{if} Z_{ik} > \delta \\ -1; \text{if} Z_{ik} > -\delta \\ 1 \text{or} -1 \text{with probability of} 0.5; \text{otherwise,} \end{cases}$$

for $k = 1, 2, 3$. We show a scatter plot of data generated in a typical simulation and the true discriminant boundary which depends on the age in Figure 2.

We computed the mean squared error (MSE) of the fitted classification boundary averaged across age for SVM_0, SVM_1 and KSVM. When a linear Mercer kernel is used and with a sample size of 500, the MSE of KSVM under sine or Gaussian coefficient function is much smaller than either SVM_0 or SVM_1: for the sine coefficient, MSE(×100)=49.8, 43.5, 6.24, respectively for SVM_0, SVM_1 and KSVM; for the Gaussian coefficient, MSE(×100)=50.7, 51.4, 2.59, respectively. This reflects the inflexibility of SVM_1 in fitting nonlinear age boundaries. When we increase the sample size to 1000, the bias in SVM_1 persists for all three scenarios. In Table 2, we summarized overall AUC and misclassification under all settings with linear kernel and Gaussian kernel. The trend in these indices is similar to the first set of simulations. That is, for more complicated functions, KSVM noticeably improves upon SVM_1 and SVM_0 with either linear or Gaussian kernel. For simpler functions such as Gaussian, using a Gaussian kernel combining age and markers improves overall performance of SVM_1.

5 Applications to two clinical studies on Huntington's disease

HD is an autosomal dominant disease caused by an expansion of CAG trinucleotide repeats at ITI5 gene on chromosome 4 (Huntington's Disease Collaborative Research Group, 1993). The disease is considered nearly fully penetrant. The inheritance of an expansion of CAG trinucleotide repeats (mutation) from a father is associated with increased penetrance to a greater extent in younger subjects than older subjects, while the effect of inheritance from a mother slightly increases over age range of their children. Majority of subjects with an expansion of CAG repeats in IT15 gene (CAG repeats ≥ 36) on one allele will develop HD if not censored by death (Kieburtz and Huntington Study Group, 1996). It is well established that the risk of HD diagnosis increases with age and CAG repeats length (Zhang et al., 2011a). A range of cognit+ive and behavioral markers may have age-varying effect on the risk of HD as well. For example, the symbol digit modality test score (SDMT, a neuropsychological measure of attention, Smith, 1982) may be more sensitive than the total motor score (a measure of motor impairment in HD, Kieburtz and Huntington Study Group, 1996) in identifying younger subjects at risk of HD while total motor score maybe more sensitive for older adults (Figure 6). In this section, various methods are applied to two large genetic epidemiological studies on HD to investigate these issues.

5.1 COHORT study results

COoperative Huntington's Observational Research Trial (COHORT, Dorsey and Huntington Study Group COHORT Investigators, 2012) is a large multi-site study that includes 42 Huntington Study Group research centers in North America and Australia. In the COHORT study, standard demographic, neurological, cognitive and behavioral instruments were administered. Individuals who met criteria for Huntington's disease (receiving a diagnostic confidence level, DCL, of 4 on the UHDRS assessment) as well as individuals at risk for HD by virtue of having a first degree relative with HD were assessed. In this example baseline data was used, and there were 338 premanifest cases and 670 controls.

Although genetic testing is available to determine whether a premanifest subject (individuals who have not been diagnosed) carries an expansion of CAG repeats, most individuals with a known family history of HD choose not to be tested since there is currently no efficacious

treatment to prevent or delay onset of disease (Williams et al., 2010). Therefore, an important research goal is to develop personalized classification to distinguish pre-symptomatic subjects who will develop HD from controls who will never develop HD without taking a genetic test. In clinical practice, HD diagnosis is based on motor symptoms, and clinicians assign a diagnostic confidence level (DCL) from UHDRS motor exam. A lower DCL category indicates lower confidence of HD, and a level of "4" indicates confirmed HD and these subjects are no longer premanifests (Paulsen et al. 2008). For a neurodegenerative disease such as HD or Alzheimer's disease (Celsis, 2000), age is one of the most important variable to control for. The goal of this analysis is to develop age-sensitive prediction to determine whether a subject who has not received a diagnosis of HD (e.g., did not receive a UHDRS DCL of 4) at the baseline visit is a pre-manifest HD case (i.e., carrying an expansion of CAG repeats, gene-positive) or a control who will not develop HD (no CAG expansion, gene-negative, will not develop HD).

To this end, we first show some descriptives of the COHORT data. In Figure 3, we present the scatter plots of a few continuous variables reported in the literature (Langbehn et al., 2007) associated with the risk of HD such as total motor score of the UHDRS (higher is more severe) and symbol digit modality test, SDMT (higher score is better). We overlay the LOWESS smoothing of the average scores in the premanifest case group and control group on the scatter plot. It is clear that none of the markers alone can discriminate the groups based on a linear boundary. We tested for nonlinearity through a regression spline model with two knots and found a significant nonlinear effect for total motor score, SDMT, and verbal uency test. It is desirable to combine markers and create nonlinear classification boundary.

We applied KSVM with Gaussian kernel to combine 19 markers in COHORT to capture the nonlinear age trend and develop an age-sensitive prediction rule. There were 6 continuous markers (e.g., body mass index (BMI), UHDRS total motor score, SDMT, verbal uency test (Mitrushina et al., 2005), and stroop test score (a weighted average of stroop color, word and interference scores, Dorsey and Huntington Study Group COHORT Investigators, 2012) and 13 binary markers (e.g., history of alcohol abuse, history of drug abuse, significant history of depression, current depression, mother affected by HD, father affected by HD). To compute an honest AUC and misclassification rate, we randomly splitted samples into a training set (n = 700, approximately 34% premanifest cases) and a testing set (n = 308) 100 times and reported the average performance indices when applying fitted model to the testing set. We compared the overall AUC and average misclassification rate over age for KSVM using all 19 markers with using a single marker for several selected markers. We compared with the penalized logistic regression with varying-coefficient age effect and accounting for interactions among markers (Paik and Hastie 2009). The varying-coefficient of age takes a nonparametric form fitted by a fourth order B-spline basis with 10 knots, and the tuning parameter was selected by five-fold cross validation. Lastly, we also compared KSVM with SVM_1 as described in section 4.

We summarize the overall sensitivity, specificity, AUC and misclassification rate using all 19 markers and several examples of using each individual marker alone in Table 3. KSVM with all 19 markers significantly improves the overall AUC (0.88) and decreases the average

misclassification rate (0.19) comparing to using a single marker alone. It is clear that combining all the markers greatly improves the prediction performance distinguishing carriers of an expansion of CAG repeats from non-carriers (controls). Among the single marker models, total motor score has the highest AUC, and the other markers have similar predictive powers that are weaker than the total motor score. The average overal AUC and sensitivity are higher than penalized logistic regression with varying coefficients and SVM_1, and the misclassification rate for KSVM is lower than these two competing methods. In Figure 4, we present a boxplot of four performance measures obtained from 100 cross validations comparing three methods to demonstrate superior performance of KSVM. The mean AUC, sensitivity and missclassification rate of KSVM are better than the other two methods, while the specificty is similar. The variability of specificity and other measures of KSVM is smaller than the competing methods, suggesting KSVM to be more robust.

In the top panels of Figure 5 we show the age-specific sensitivity and specificity. We see a decreasing age trend in sensitivity which suggests it is easier to screen presymptomatic cases from the population for younger subjects than for older subjects, i.e., the predictive score is more sensitive for younger subjects. When a subject shows subtle motor signs or cognitive decline at an early age, it is an indication of increased likelihood of developing HD in the future since such signs may be rarely present in controls of similar age. When a subject shows signs of clinical symptoms at an older age, however, it is less predicative of HD disease status since controls at older age may also show similar signs.

Combining all markers significantly improves over using single marker. For example, total motor score and SDMT have sensitivities decreasing to zero for older ages (non-age-corrected raw SDMT was used). We show the specificity in the upper right panel of Figure 5. As expected, specificity increases with age, which suggests it is easier to screen controls from the sample for older subjects. When the clinical markers are absent by an old age, it is more likely a subject will never develop the disease, and therefore the score is more specific for older subjects. Furthermore, since a subject at-risk for HD is mostly likely to develop HD between age 30 and 50 (Foroud et al., 1999), the increasing trend in specificity is consistent with the clinical observation that an older subject who does not develop HD by a certain age is more likely to be in the control group. When compared to the penalized logistic regression, we see an improvement in sensitivity especially at the younger age.

In the bottom panels of Figure 5, we show trajectories of the age-specific AUC and misclassification rate. Again, we see at each age, using multiple markers has superior performance than using each single marker. The general trend shows that considering both sensitivity and specificity, it is easier to predict the risk status of HD in an older subject than a younger subject since the AUC increases with age and the misclassification status decreases with age. We can also see from the figure that the combined predictive score maybe more accurate in the older age range, for example, the AUC>0.85 for subjects with age> 38. When splitting samples by the median age (47), the AUC is 0.84 for younger subjects and 0.89 for older subjects. The AUC of the KSVM is higher than the logistic regression from age 20 to 55, and similar from 55 to 70. Same trend is observed for the misclassification rate.

To further investigate the relative ranking of markers, the first two subfigures in Figure 6 present the age-specific predictive effect of several markers from age 20 to 70. These effects are computed as differences in the fitted discriminant functions between values 1 and 0 of a particular binary marker or as differences of 1/4 standard deviation units increase of a particular continuous marker with other markers fixed at sample means in the local age window (5-year). It shows the markers expressing different trends: some with increasing age effect (seeing a mental health professional) and decreasing effect (father's HD status). More importantly, we see that the relative magnitude of the marker effect changes across age and the ranking of the importance of markers based on the magnitude of shifts in their classification function also varies with age. For example, SDMT score is more important than the total motor score for younger subjects, while the total motor score dominates other markers for older subjects (age 45 or above).

In summary, this analysis show that markers' sensitivity and specificity vary in predicting at risk for HD according to age. Combining informative markers significantly improves prediction accuracy. The most important marker for younger subjects is SDMT while it is total motor score for older subjects.

5.2 PREDICT-HD study results

We illustrate our methods through a second example, PREDICT-HD (Paulsen et al., 2008), a 32-site observational study of HD focusing on premanifest subjects followed from the prodromal phase through to post-diagnosis. To date, the main study has 1314 total participants, 1013 of whom were gene-expanded cases and 301 of whom were non-expanded controls. The individual follow up period spans 10 years with annual or biennial measurements on variables in important domains of motor, cognitive, psychiatric as well as brain imaging. The number of subjects at each visit ranges from 43 to 380. One of the major goals of PREDICT-HD is to discover markers for predicting onset of HD diagnosis based on motor symptoms in a short study period in premanifests subjects. Such information is valuable for planning recruit of a future clinical trial on HD. Thus, here our outcome of interest is the risk of a pre-symptomatic subject at baseline receiving HD diagnosis during the study period. That is, to predict risk of conversion: risk of a subject with DCL<4 (no diagnosis) at the baseline converting to DCL=4 (receive a confirmed clinical diagnosis) in the study period. This outcome of interest in this section is conversion status distinguishes PREDICT analysis from COHORT analysis in the previous section (outcome mutation carrier status).

Our analysis included a subsample of 671 gene-expanded cases from PREDICT-HD study who were not diagnosed with HD at the baseline. There were 107 converters who received a disease diagnosis during the study period. Five markers (gender, CAG repeats, total motor score, TFC and stroop color score) were used to predict the age-specific conversion status in the age range from 25 to 65. We applied both KSVM (with Gaussian kernel) and penalized logistic regression (Paik and Hastie 2009) with nonparametric varying coefficient (B-spline basis expansion with 10 knots) to the data for comparison similar to the COHORT study. The tuning parameter was selected by five-fold cross-validation.

We show some descriptives of the markers included in the analyses in the top panels of Figure 7. We present the scatter plots of baseline total motor score and stroop color score with overlaid LOWESS plots as examples. Although the figure hints the mean total motor score to be different in converters and non-converters, a linear separation boundary does not appear to be adequate. Similar pattern can be seen for the stroop score. We therefore combine all five markers to perform classification with a nonlinear boundary. The bottom panels of Figure 7 show the results. From bottom left subfigure, we see that the age-specific sensitivity of KSVM is much higher compared to penalized logistic regression in the younger age range (before 43 years old). The specificity of the two methods is similar (results not shown). For the older age range, their performance is similar. The right panel shows the standardized effects of four continues markers (measured in 1/4 standard deviation unit of each marker). Baseline total motor score has the largest effect across all age range, suggesting the importance of this marker in tracking disease progression. Among the other markers, total functional capacity has larger effect for younger subjects (less than age 37), while these markers have similar magnitude of effect for older age range.

In summary, this analysis shows that KSVM creates much more sensitive predictive score especially for younger subjects. In predicting conversion status during a fixed time period, baseline total motor score has dominating effect over other markers.

6 Discussion

We have proposed a local smoothing classification method to predict disease risk accounting for its age-dependent effect. Age has clear clinical interpretation and represents a constellation of underlying unobserved biological and physiological factors. Constructing age-specific prediction rules facilitates studying timing of intervention and discovering markers useful to guide personalized treatments. The fitted coefficients $\beta(w)$ depict age-sensitive profiles of the markers on disease risk. Furthermore, the obtained age-dependent predictive scores can be used to allocate patients into risk groups. Therefore the developed methods can be used to recruit high-risk patients for clinical trials based on a subject's age and marker values to improve efficiency of the trial. In the application example, we classified HD premanifest case/control status for presymptomatic individuals where all subjects with CAG≥ 36 belong to the case group (they will develop HD a future time point). It would be interesting to use the actual CAG repeat length in a future work and to classify more refined groups of cases (e.g., close or far to disease onset). It may also be desirable to examine predictive powers of other markers such as brain imaging measures in a future analysis.

Here we considered markers with age-dependent effects, but it is easy to incorporate markers with constant effects. For example, an iterative backfitting procedure can be used to include markers $\boldsymbol{Z}$ with age-invariant effects and fit decision boundaries such as

$$\alpha(w)+\boldsymbol{X}^T\boldsymbol{\beta}(w)+\boldsymbol{Z}^T\boldsymbol{\gamma}.$$

Specifically, at a given γ, $\alpha(w)$ and $\beta(w)$ will be fitted through the developed approaches. Then fixing these functions at their fitted values, an update of γ is obtained through a regular

SVM procedure without smoothing. These two steps will be iterated until convergence. We can extend the current approach when there is an additional marker that needs special attention (e.g., BMI or CAG repeats length). We can then extend our method to incorporate a two-dimensional coefficient function, i.e. $\beta(w, u)$, and apply two-dimensional local kernel smoothing. It is also easy to extend the current methods to multi-category outcomes and to continuous outcomes.

Large margin classification with other penalty functions are discussed in Zhu et al. (2003) (i.e., 1-norm SVM) and Zou and Yuan (2008) (i.e., F_∞-norm SVM). We have not considered marker selection in the current local smoothing setting. It may be possible to use some of the other penalty functions to perform marker selection so that the marker without any effect at the entire range of age will be automatically excluded. We do not discuss effective handling of correlated markers here. Lastly, our simulation results show that different choices of Mercer kernel may lead to slight difference in prediction accuracy. A procedure that maximizes performance over a class of Mercer kernels is conceivable. These topics worth some future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by NIH grants NS073671, NS036630, P01CA142538, National Center for Research Resources grant UL1 RR025747, Huntington's Disease Society of America, and Cure Huntington Disease Initiative, Family Fund, NIH data repository, PREDICT-HD study investigators, and COHORT study investigators. The authors wish to thank the Associate Editor and two anonymous reviewers for their constructive and helpful comments and suggestions on this paper.

References

Cai Z, Fan J, Li R. Efficient estimation and inferences for varying coefficient models. Journal of the American Statistical Association. 2000; 95:888–902.

Celsis P. Age-related cognitive decline, mild cognitive impairment or preclinical alzheimer's disease? Annals of Medicine. 2000; 32:6–14. [PubMed: 10711572]

Dorsey ER. Huntington Study Group COHORT Investigators. Characterization of a large group of individuals with huntington disease and their relatives enrolled in the cohort study. PLoS ONE. 2012; 7 (2, Article ID e29522).

Foroud T, Gray J, Ivashina J, Conneally PM. Differences in duration of huntingtons disease based on age at onset. Journal of Neurology, Neurosurgery & Psychiatry. 1999; 66:52–56.

Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on huntingtons disease chromosomes. Cell. 1993; 72:971–983. [PubMed: 8458085]

Kieburtz K. Huntington Study Group. The unified huntington's disease rating scale: reliability and consistency. Movement Disorders. 1996; 11:136–142. [PubMed: 8684382]

Ladicky L, Torr PHS. Locally linear support vector machines. ICML2011. 2011:985–992.

Langbehn DR, Paulsen JS. Huntington Study Group. Predictors of diagnosis in huntington disease. Neurology. 2007; 68(20):1710–1717. [PubMed: 17502553]

Lin Y. Support vector machines and the bayes rule in classification. Data Mining and Knowledge Discovery. 2002; 6:259–275.

Mitrushina, M.; Boone, KB.; Razani, J.; D'Elia, LF. Handbook of normative data for neuropsychological assessment. New York: Oxford University Press; 2005.

Moguerza JM, Munoz A. Support vector machines with applications. Statistical Science. 2006; 21(3): 322–336.

Orru G, Pettersson-Yeoa W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. Neuroscience and Biobehavioral Reviews. 2012; 36(4):1140–1152. [PubMed: 22305994]

Paulsen J, Hayden M, Stout JC, Langbehn DR, Aylward E, Ross CA, Guttman M, Nance M, Kieburtz K, Oakes D, Shoulson I, Kayson E, Johnson S, Penziner E. Predict-HD Investigators of the Huntington Study Group. Preparing for preventive clinical trials: the predict-hd study. Archives of Neurology. 2006; 65(6):883–890. [PubMed: 16769871]

Paulsen JS, Langbehn DR, Stout JC, Aylward E, Ross CA, Nance M, Guttman M, Johnson S, MacDonald M, Beglinger LJ, Duff K, Kayson E, Biglan K, Shoulson I, Oakes D, Hayden M. Detection of huntington's disease decades before diagnosis: the predict-hd study. Journal of Neurology, Neurosurgery & Psychiatry. 2008; 79:874–880.

Pepe MS, Cai T, Longton G. Combining predictors for classification using the area under the receiver operating characteristic curve. Biometrics. 2006; 62:221–229. [PubMed: 16542249]

Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. American Journal of Epidemiology. 2004; 159:882–890. [PubMed: 15105181]

Shen X, Tseng GC, Zhang X, Wong WH. On ψ-learning. Journal of the American Statistical Association. 2003; 98:724–734.

Smith, A. Symbol digit modalities test: Manual. Los Angeles: Western Psychological Services; 1982.

Steinwart I, Scovel C. Fast rates for support vector machines using gaussian kernels. Annals of Statistics. 2007; 35:575–607.

Vapnik, V. The Nature of Statistical Learning Theory. New York: Springer-Verlag; 1995.

Wahba, G. Spline Models for Observational Data. SIAM; 1990.

Wang J, Shen X, Pan W. On efficient large margin semisupervised learning: Method and theory. Journal of Machine Learning Research. 2009a; 10:719–742. [PubMed: 24678270]

Wang L, Kai B, Li R. Local rank inference for varying coefficient models. Journal of the American Statistical Association. 2009b; 104(488):1631C1645. [PubMed: 20657760]

Ware JH. The limitations of risk factors as prognostic tools. The New England Journal of Medicine. 2006; 355(4):2615–2617. [PubMed: 17182986]

Wei Z, Wang K, Qu H-Q, Zhang H, Bradfield J, Kim C, Frackleton E, Hou C, Glessner JT, Chiavacci R, Stanley C, Monos D, Grant SFA, Polychronakos C, Hakonarson H. From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. PLoS Genetics. 2009; 5(10):e1000678. [PubMed: 19816555]

Williams JK, Erwin C, Juhl A, Mills J, Brossman B, Paulsen JS. and the I-RESPOND-HD Investigators of the Huntington Study Group. Personal factors associated with reported benefits of huntington disease family history or genetic testing. Genetic Testing and Molecular Biomarkers. 2010; 14(5):629–636. [PubMed: 20722493]

Wu Y, Liu Y. Functional robust support vector machines for sparse and irregular longitudinal data. Journal of Computational and Graphical Statistics. 2012

Zhang HH, Ahn J, Lin X, Park C. Gene selection using support vector machine with non-convex penalty. Bioinformatics. 2006; 22(1):88–95. [PubMed: 16249260]

Zhang Y, Long JD, Mills JA, Warner JH, Lu W, Paulsen JS. and the PREDICT-HD Investigators and Coordinators of the Huntington Study Group. Indexing disease progression at study entry with individuals at-risk for huntington disease. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2011a; 156B(7):751–763.

Zhang Z, Ladicky L, Torr PHS, Saffari A. Learning anchor planes for classification. NIPS. 2011b

Zhu, J.; Rosset, S.; Hastie, T.; Tibshirani, R. Neural Information Processing Systems. MIT Press; 2003. 1-norm support vector machines; p. 16

Zou H, Yuan M. The f_{∞}-norm support vector machine. Statistica Sinica. 2008; 18(1):379–398.

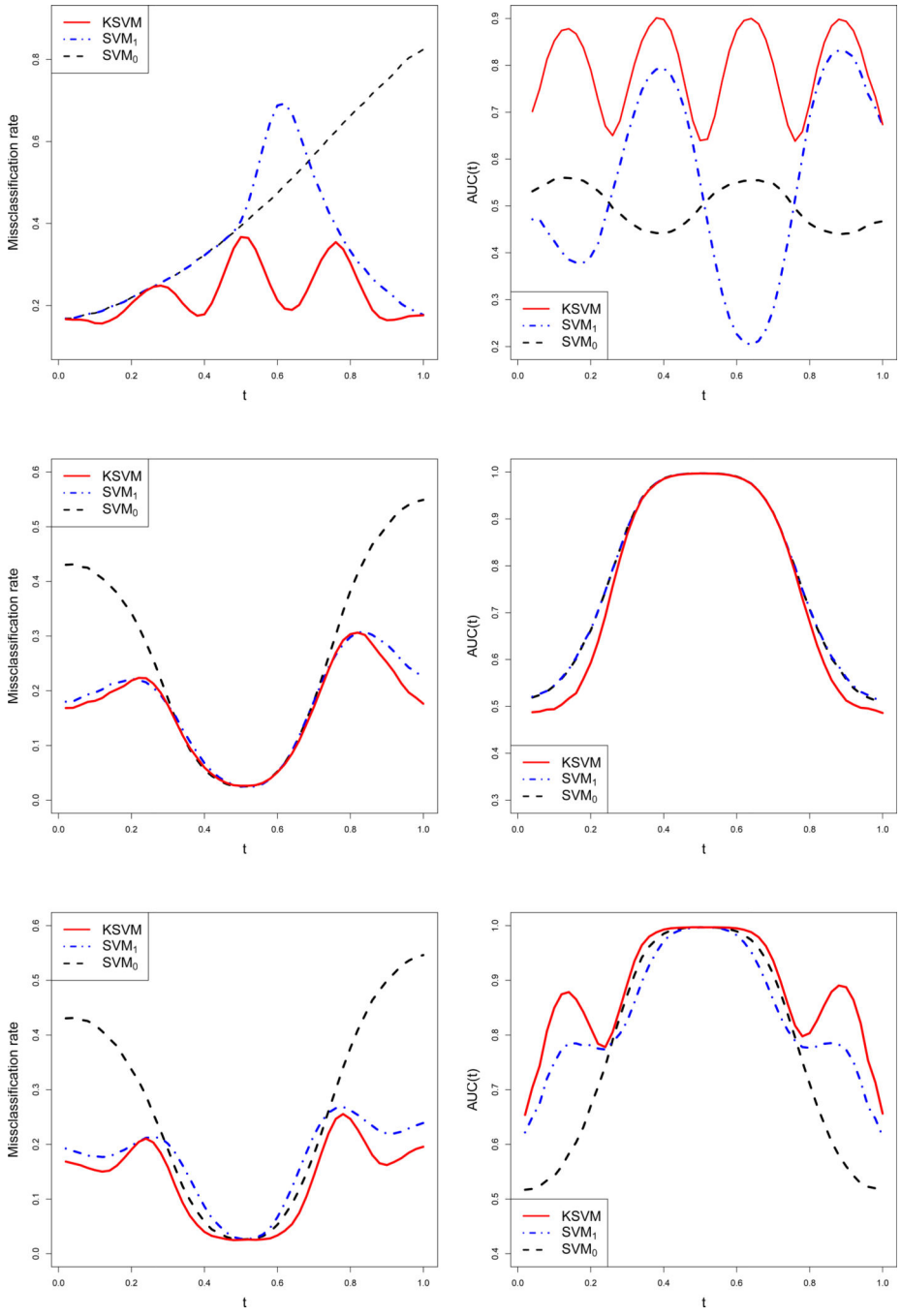

Figure 1.

(Simulation 1) Age-specific misclassification rate (left) and AUC (right) for SVM_0, SVM_1 and KSVM. The corresponding analysis from the top to the bottom are: using X_{i1}, using X_{i2} and using multiple markers.

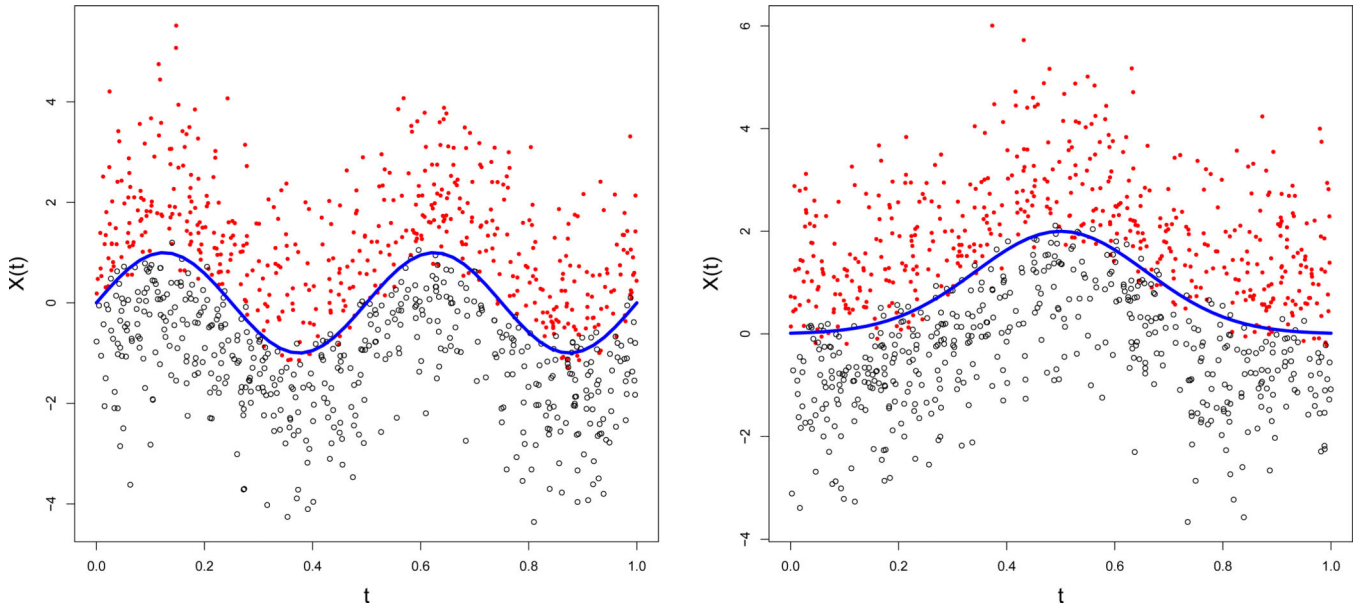

Figure 2.
(Simulation 2) True classification boundary and a typical set of simulated data.

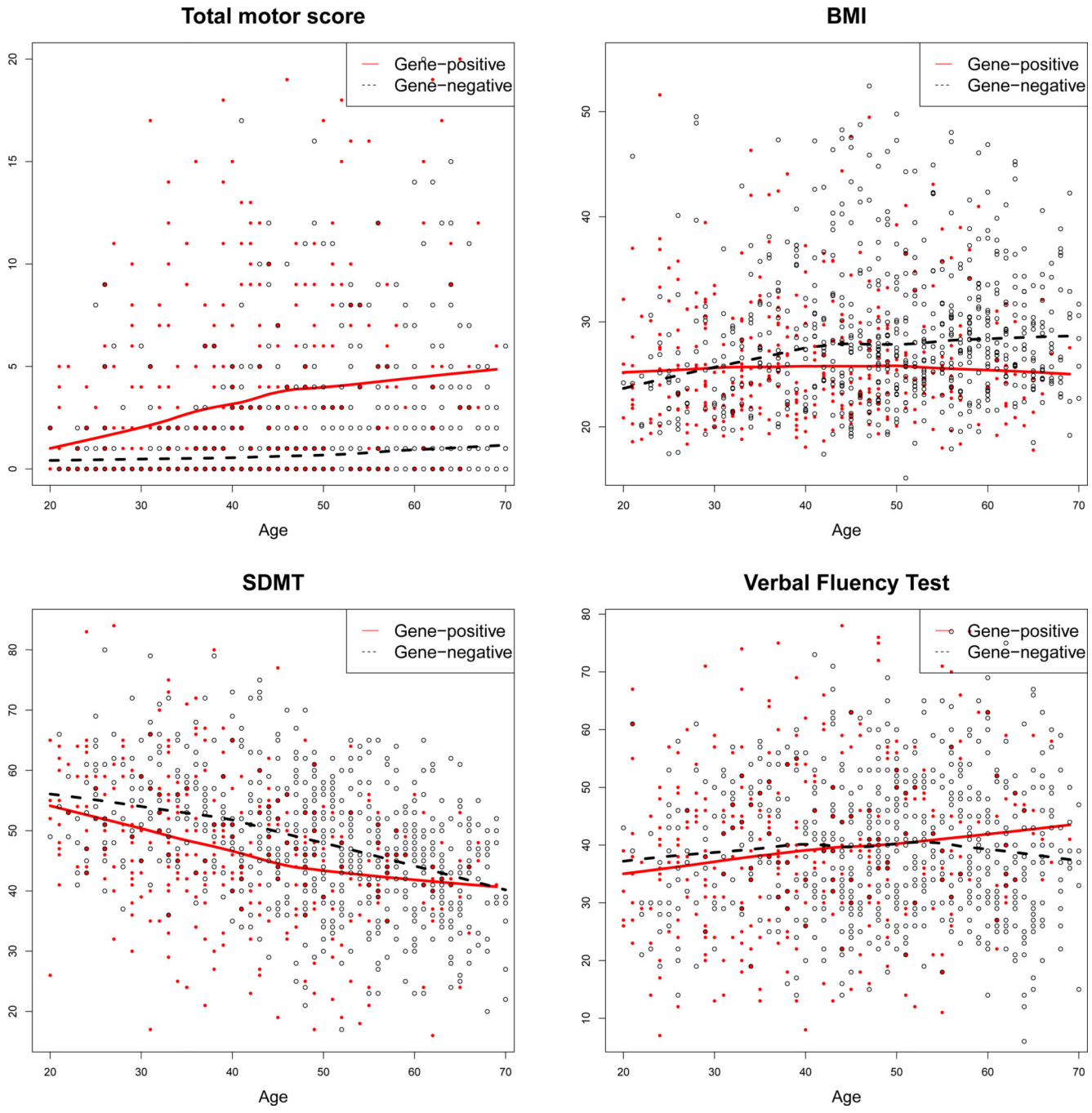

Figure 3.

Descriptive scatter plots of several continuous markers and lowess smoothed mean curves in COHORT

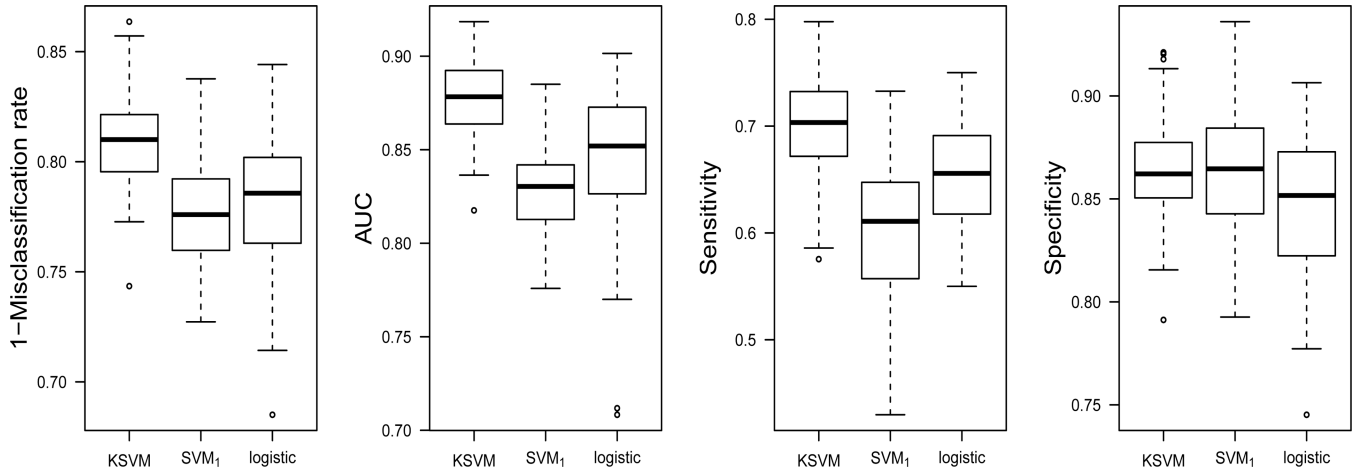

Figure 4.

Comparison of KSVM, SVM_1 and penalized logistic using 19 markers in predicting at-risk status of Huntington's disease with COHORT premanifest subjects (overall 1-Misclassification Rate, AUC, Sensitivity, and Specificity).

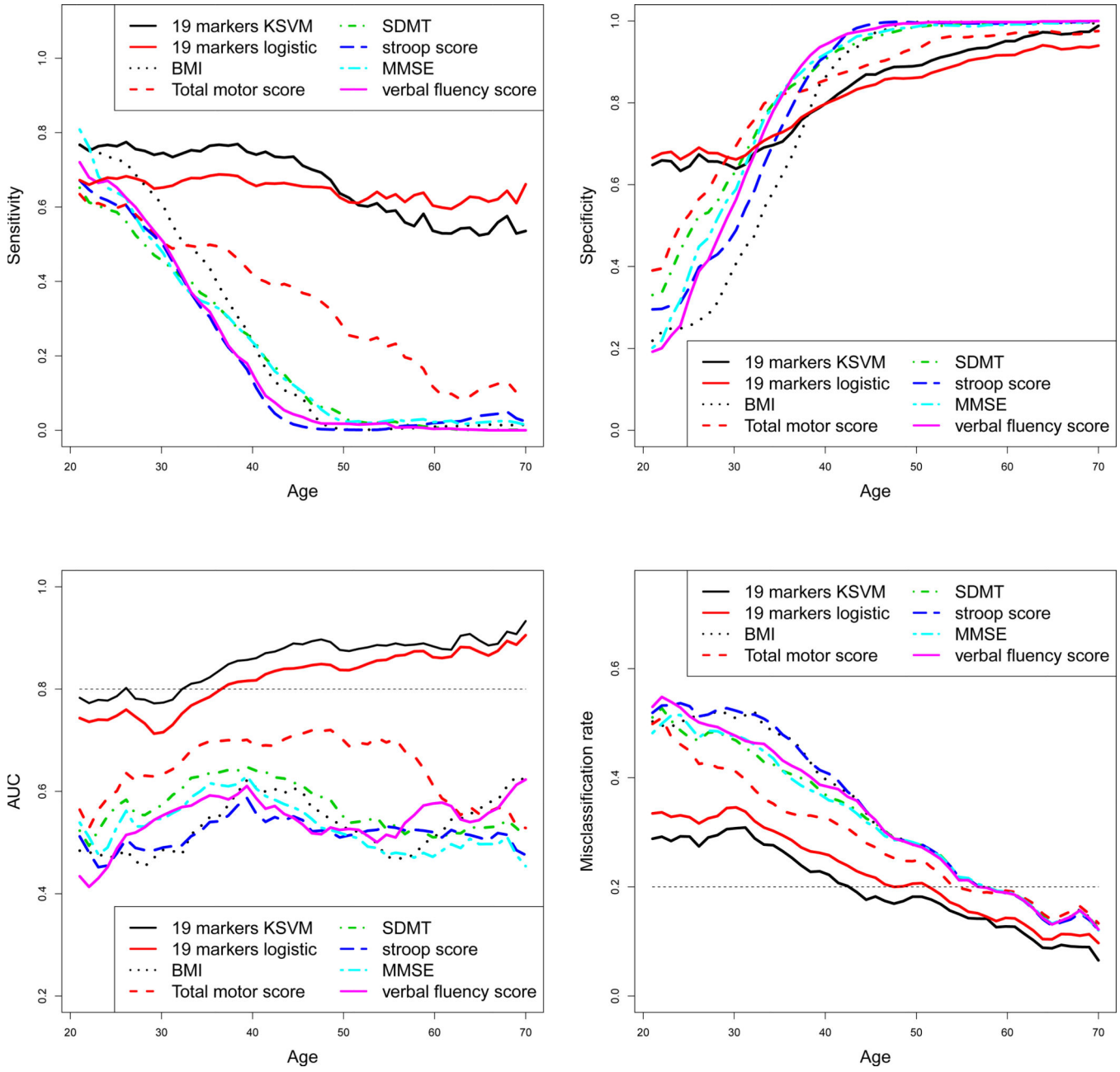

Figure 5.

Comparison of 19-marker penalized logistic, 19-marker KSVM and single-marker KSVM in predicting at-risk status of Huntington's disease with COHORT premanifest subjects (age-specific sensitivity, specificity, AUC and misclassification rate).

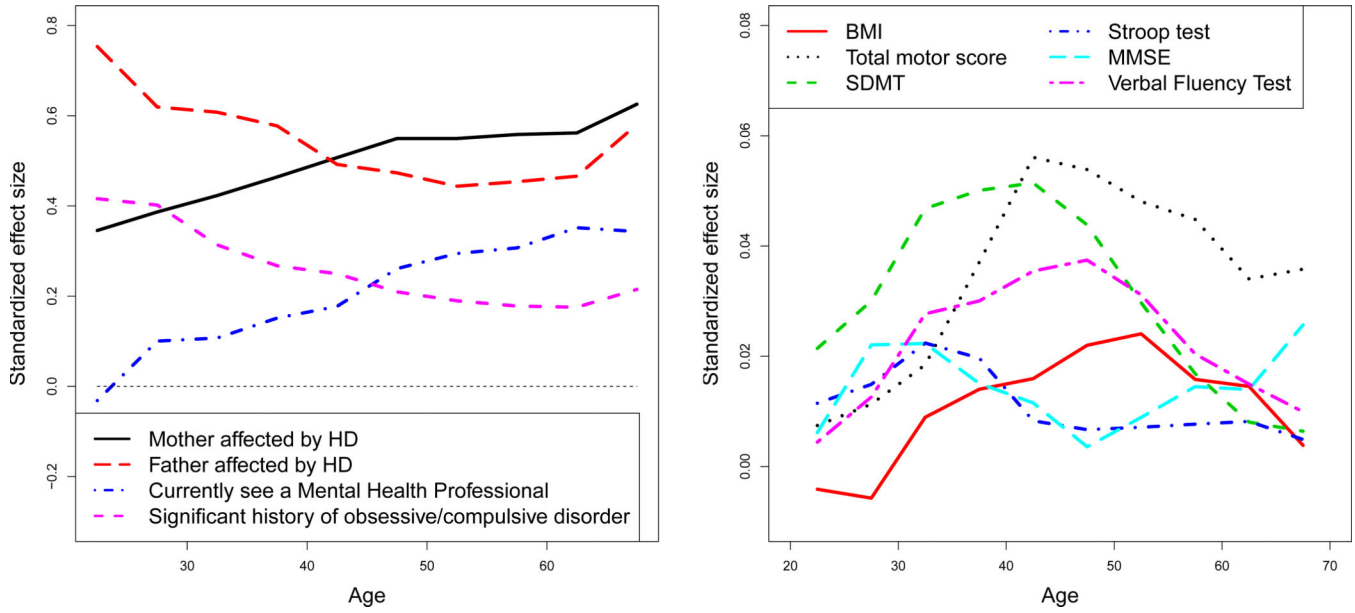

Figure 6.
Standardized effect for key markers in COHORT study fitted by KSVM.

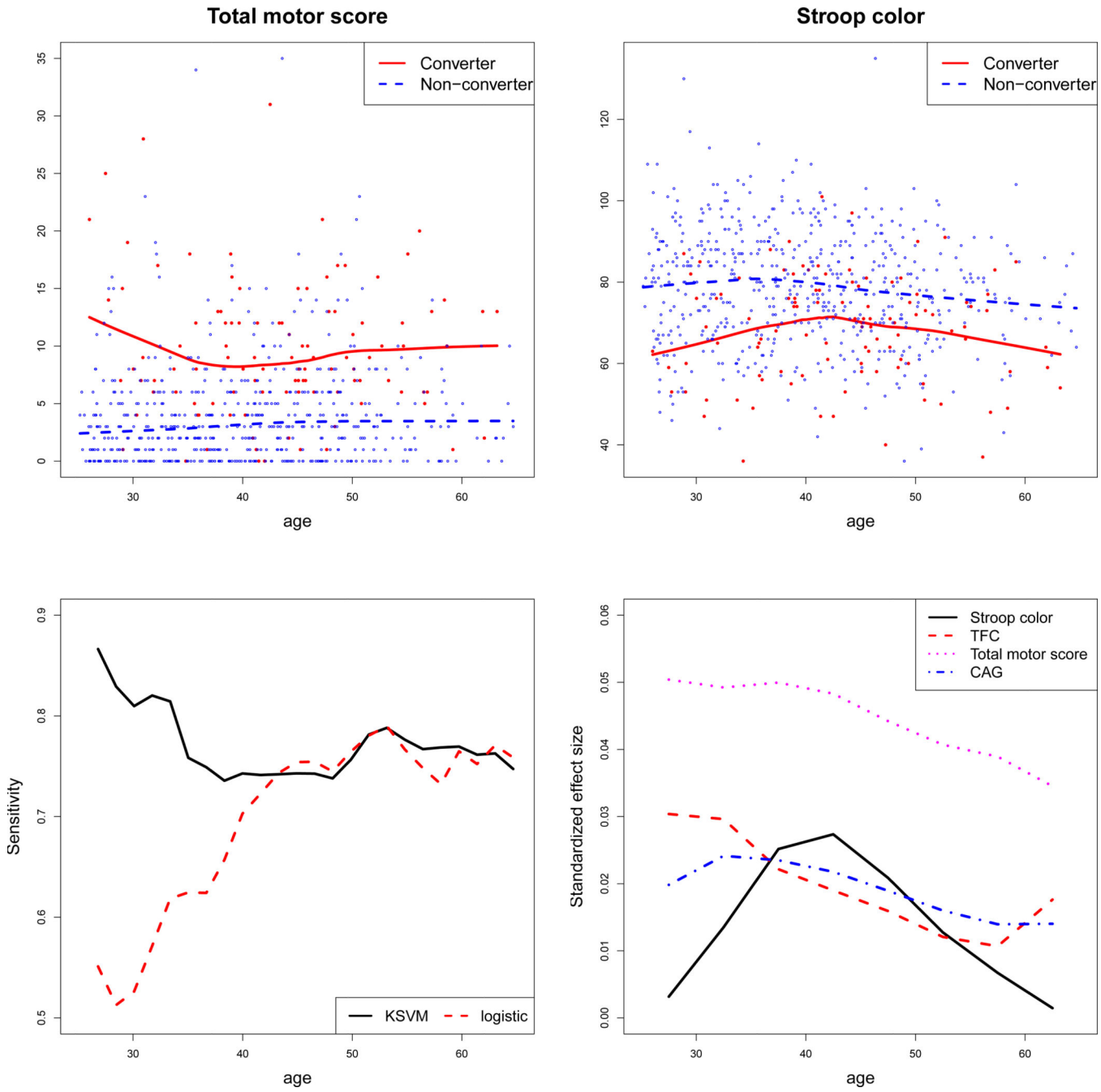

Figure 7.

Age-specific descriptives, sensitivity, and standardized effect for predicting HD conversion status in PREDICT-HD premanifest subjects.

Table 1

Summary of simulation results from Simulation 1 (retrospective data generation)

Marker used	Mercer kernel	Index	SVM_0†	SVM_1‡	KSVM§	SVM_0†	SVM_1‡	KSVM§
			***n*=500**			***n*=1000**		
X_{i1}	Linear	Miss*	0.439	0.334	0.240	0.440	0.334	0.228
		AUC**	0.500	0.738	0.833	0.498	0.743	0.848
	Gaussian	Miss	0.446	0.277	0.233	0.442	0.260	0.223
		AUC	0.500	0.760	0.825	0.501	0.783	0.838
X_{i2}	Linear	Miss	0.262	0.174	0.164	0.264	0.170	0.161
		AUC	0.819	0.884	0.899	0.818	0.885	0.903
	Gaussian	Miss	0.263	0.165	0.166	0.265	0.161	0.161
		AUC	0.782	0.899	0.890	0.780	0.902	0.897
Multiple	Linear	Miss	0.267	0.174	0.143	0.265	0.168	0.133
		AUC	0.813	0.897	0.932	0.817	0.902	0.941
	Gaussian	Miss	0.270	0.172	0.151	0.265	0.158	0.140
		AUC	0.792	0.896	0.915	0.799	0.909	0.925

*Overall misclassification rate averaged over age;

**Overall AUC averaged over age;

†Ignoring age effect;

‡A parametric linear age effect;

§Local smoothing of age effect.

Table 2

Summary of simulation results from Simulation 2 (prospective data generation)

Marker used	Mercer kernel	Index	SVM_0†	SVM_1‡	KSVM§	SVM_0†	SVM_1‡	KSVM§
			***n*=500**			***n*=1000**		
X_{i1}	Linear	Miss†	0.168	0.153	0.084	0.169	0.153	0.081
		AUC‡	0.930	0.937	0.979	0.930	0.937	0.981
	Gaussian	Miss	0.170	0.153	0.087	0.170	0.152	0.082
		AUC	0.903	0.922	0.972	0.902	0.928	0.978
X_{i2}	Linear	Miss	0.164	0.164	0.079	0.166	0.166	0.080
		AUC	0.931	0.930	0.984	0.930	0.929	0.985
	Gaussian	Miss	0.164	0.085	0.083	0.163	0.081	0.080
		AUC	0.897	0.980	0.978	0.900	0.983	0.982
Multiple	Linear	Miss	0.150	0.146	0.084	0.150	0.143	0.081
		AUC	0.935	0.940	0.977	0.936	0.944	0.978
	Gaussian	Miss	0.153	0.138	0.095	0.152	0.120	0.081
		AUC	0.920	0.944	0.972	0.921	0.957	0.979

*Overall misclassification rate averaged over age;

**Overall AUC averaged over age;

†Ignoring age effect;

‡A parametric linear age effect;

§Local smoothing of age effect.

Table 3

Overall performance over age for multiple markers models compared with various single marker models.

	Misclassification	AUC	Sensitivity	Specificity
All markers (KSVM)	0.190 (0.020)‡	0.878 (0.020)	0.700 (0.046)	0.864 (0.024)
All markers (SVM_1)	0.223 (0.023)	0.829 (0.024)	0.604 (0.062)	0.864 (0.032)
All markers (Penalized logistic regression†)	0.218 (0.029)	0.844 (0.038)	0.655 (0.048)	0.846 (0.032)
Total Motor Score	0.276 (0.021)	0.731 (0.034)	0.403 (0.082)	0.885 (0.038)
SDMT	0.307 (0.024)	0.668 (0.037)	0.258 (0.068)	0.910 (0.038)
BMI	0.322 (0.023)	0.657 (0.028)	0.296 (0.125)	0.870 (0.058)
Mini-Mental Exam	0.306 (0.022)	0.647 (0.033)	0.272 (0.061)	0.904 (0.033)
Verbal Fluency Test	0.314 (0.021)	0.651 (0.031)	0.242 (0.069)	0.907 (0.039)
Stroop score	0.326 (0.020)	0.639 (0.034)	0.226 (0.096)	0.898 (0.049)
Father affected by HD	0.293 (0.022)	–	0.366 (0.070)	0.880 (0.043)
Mother affected by HD	0.321 (0.023)	–	0.326 (0.121)	0.859 (0.070)
Currently see a Mental Health Professional	0.319 (0.024)	–	0.277 (0.104)	0.886 (0.052)
Significant history of depression	0.322 (0.023)	–	0.256 (0.099)	0.893 (0.053)
History of alcohol abuse	0.328 (0.025)	–	0.236 (0.112)	0.894 (0.061)
Significant history of suicidal ideation	0.332 (0.023)	–	0.239 (0.124)	0.887 (0.069)
History of tobacco abuse	0.330 (0.022)	–	0.240 (0.120)	0.890 (0.062)
Significant history of OCD	0.322 (0.024)	–	0.231 (0.097)	0.906 (0.050)
History of drug abuse	0.328 (0.022)	–	0.223 (0.103)	0.901 (0.058)
Current depression	0.326 (0.022)	–	0.180 (0.103)	0.925 (0.059)

† Proposed in Paik and Hastie (2009).

‡ Mean and empirical standard deviation for 100 cross validations.